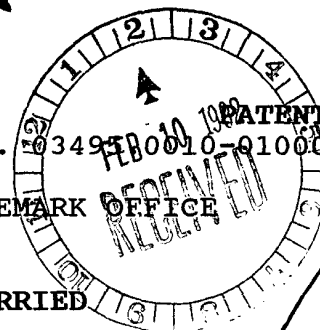




Attorney Docket No. 03495-00010-01000



#15
6/77
2/12/92

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

) HANDCARRIED
)
)
)
)

MARC ALIZON et al.

Serial No.: 158,652

) Group Art Unit: 1805
)
)

Filed: February 22, 1988

) Examiner: J. Railey
)
)

For: CLONED DNA SEQUENCES RELATED TO THE GENOMIC RNA
OF LYMPHADENOPATHY ASSOCIATED VIRUS (LAV) AND
PROTEINS ENCODED BY SAID LAV GENOMIC RNA

Honorable Commissioner of Patents
and Trademarks
Washington, D.C. 20231

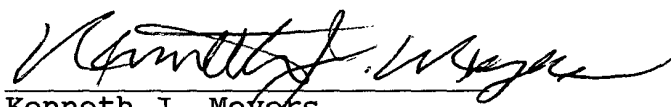
Sir:

SUBMISSION OF COPY OF PRIORITY DOCUMENT

Further to the Examiner's telephone request, enclosed
herewith is a copy of United Kingdom Priority Document
No. 8429099.

Respectfully submitted,

FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER

By: 
Kenneth J. Meyers
Reg. No. 25,146

Dated: February 6, 1992

19/11/84 B3662 FAT*** 10.00

The Comptroller
The Patent Office
25 Southampton Buildings
London, WC2A 1AY

Fee: £10.00

REQUEST FOR GRANT OF A PATENT

THE GRANT OF A PATENT IS REQUESTED BY THE UNDERSIGNED ON THE BASIS OF THE PRESENT APPLICATION

1 Agent's Reference JJD/EEF/26804

11 Title of Invention CLONED DNA SEQUENCES RELATED TO THE GENOMIC RNA OF
LYMPHADENOPATHY-ASSOCIATED VIRUS (LAV) AND PROTEINS
ENCODED BY SAID LAV GENOMIC RNA.

III Applicant or Applicants (See note 2)

Name (First or only applicant) INSTITUT PASTEUR

Country FRANCE State ADP Code No.

Address 25-28 Rue du Dr. Roux,
75724 Paris Cedex 15, France.

Name (of second applicant, if more than one) Centre National
de la Recherche Scientifique Country FRANCE State

Address 15 Quay Anatole France
75007 Paris, France.

IV Inventor (see note 3)

or

(b) A statement on Patents Form No. 7/77 is/will be furnished

V Name of Agent (if any) (See note 4)

Reddie & Grose

ADP CODE NO

VI Address for Service (See note 5)

16 Theobalds Road
London WC1X 8PL

VII Declaration of Priority (See note 6)

Country

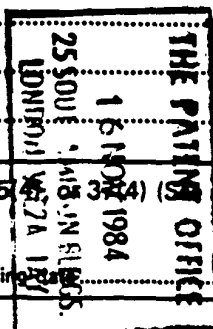
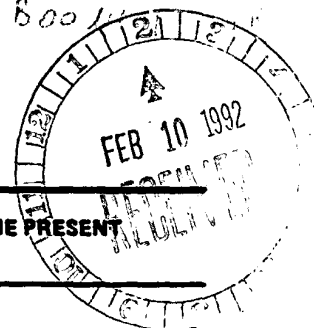
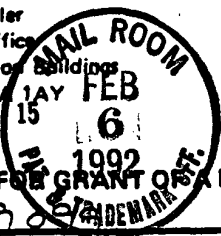
Filing date

File number

VIII The Application claims an earlier date under Section 8(3), 12(6), 15(4), or 35(4) (See note 7)

Section No.

Earlier application or patent number and filing date



IX Check List (To be filled in by applicant or agent)

- | | |
|---|---|
| <p>A The application contains the following number of sheet(s)</p> <p>1 Request 1 Sheet(s)</p> <p>2 Description 17 Sheet(s)</p> <p>3 Claim(s) 2 Sheet(s)</p> <p>4 Drawing(s) 26 Sheet(s)</p> <p>5 Abstract 0 Sheet(s)</p> | <p>B The application as filed is accompanied by:-</p> <p>1 Priority document No</p> <p>2 Translation of priority document No</p> <p>3 Request for Search No</p> <p>4 Statement of Inventorship and Right to Apply No</p> <p>5</p> |
|---|---|

X It is suggested that Figure No 1 of the drawings (if any) should accompany the abstract when published.

XI Signature (See note 8)


Reddie & Grose, Agents for the Applicant(s)

NOTES:

1. This form, when completed, should be brought or sent to the Patent Office together with the prescribed fee and two copies of the description of the invention, and of any drawings.
2. Enter the name and address of each applicant. Names of individuals should be indicated in full and the surname or family name should be underlined. The names of all partners in a firm must be given in full. Bodies corporate should be designated by their corporate name and the country of incorporation and, where appropriate, the state of incorporation within that country should be entered where provided. Full corporate details, eg "a corporation organised and existing under the laws of the State of Delaware, United States of America," trading styles, eg "trading as xyz company", nationality, and former names, eg "formerly [known as] ABC Ltd." are not required and should not be given. Also enter applicant(s) ADP Code No. (if known).
3. Where the applicant or applicants is/are the sole inventor or the joint inventors, the declaration (a) to that effect at IV should be completed, and the alternative statement (b) deleted. If, however, this is not the case the declaration (a) should be struck out and a statement will then be required to be filed upon Patent Form No 7/77.
4. If the applicant has appointed an agent to act on his behalf, the agent's name and the address of his place of business should be indicated in the spaces available at V and VI. Also insert agent's ADP Code No. (if known) in the box provided.
5. An address for service in the United Kingdom to which all documents may be sent must be stated at VI. It is recommended that a telephone number be provided if an agent is not appointed.
6. The declaration of priority at VII should state the date of the previous filing and the country in which it was made and indicate the file number, if available.
7. When an application is made by virtue of section 8(3), 12(6), 15(4), or 37(4) the appropriate section should be identified at VIII and the number of the earlier application or any patent granted thereon identified.
8. Attention is directed to rules 90 and 106 of the Patent Rules 1982.
9. Attention of applicants is drawn to the desirability of avoiding publication of inventions relating to any article, material or device intended or adapted for use in war (Official Secrets Acts, 1911 and 1920). In addition after an application for a patent has been filed at the Patent Office the comptroller will consider whether publication or communication of the invention should be prohibited or restricted under section 22 of the Act and will inform the applicant if such prohibition is necessary.
10. Applicants resident in the United Kingdom are also reminded that, under the provisions of section 23 applications may not be filed abroad without written permission or unless an application has been filed not less than six weeks previously in the United Kingdom for a patent for the same invention and no direction prohibiting publication or communication has been given or any such direction has been received.

Cloned DNA sequences related to the genomic RNA of lymphadenopathy-associated-virus (LAV) and proteins encoded by said LAV genomic RNA

The invention relates to cloned DNA sequences
5 indistinguishable from genomic RNA and DNA of lymphadenopathy-associated virus (LAV), a process for their preparation and their uses. It relates more particularly to stable probes including a DNA sequence which can be used for the detection of the LAV virus or related viruses
10 or DNA proviruses in any medium, particularly biological samples containing any of them. The invention also relates to polypeptides, whether glycosylated or not, encoded by said DNA sequences.

Lymphadenopathy-associated virus (LAV) is a human
15 retrovirus first isolated from the lymph node of a homosexual patient with lymphadenopathy syndrome, frequently a prodrome or a benign form of acquired immune deficiency syndrome (AIDS). Subsequently other LAV isolates have been recovered from patients with AIDS or pre-AIDS. All available data are consistent with the virus being the causative
20 agent of AIDS.

A method for cloning such DNA sequences has already been disclosed in British Patent Application Nr.
84 23859 filed on September 19, 1984. Reference is here-
25 after made to that application as concerns subject matter in common with the further improvements to the invention disclosed herein.

The present invention aims at providing additional new means which should not only also be useful for the
30 detection of LAV or related viruses (hereafter more generally referred to as "LAV viruses"), but also have more versatility, particularly in detecting specific parts of the genomic DNA of said viruses whose expression products are not always directly detectable by immunological
35 methods.

The present invention further aims at providing

polypeptides containing sequences in common with polypeptides encoded by the LAV genomic RNA. It relates even more particularly to polypeptides comprising antigenic determinants included in the proteins encoded and expressed by the LAV genome occurring in nature. An additional object of the invention is to further provide means for the detection of proteins related to LAV virus, particularly for the diagnosis of AIDS or pre-AIDS or, to the contrary, for the detection of antibodies against the LAV virus or proteins related therewith, particularly in patients afflicted with AIDS or pre-AIDS or more generally in asymptomatic carriers and in blood-related products. Finally the invention also aims at providing immunogenic polypeptides, and more particularly protective polypeptides for use in the preparation of vaccine compositions against AIDS or related syndroms.

The present invention relates to additional DNA fragments, hybridizable with the genomic RNA of LAV as they will be disclosed hereafter, as well as with additional cDNA variants corresponding to the whole genomes of LAV viruses. It further relates to DNA recombinants containing said DNAs or cDNA fragments.

The invention relates more particularly to a cDNA variant corresponding to the whole of LAV retroviral genomes, which is characterized by a series of restriction sites in the order hereafter (from the 5' end to the 3' end).

The coordinates of the successive sites of the whole LAV genome (restriction map) are indicated hereafter too, with respect to the Hind III site (selected as of coordinate 1) which is located in the R region. The coordinates are estimated with an accuracy of ± 200 bp :

	Hind III	0
	Sac I	50
35	Hind III	520
	Pst I	800
	Hind III	1 100

	Bgl II	1 500
	Kpn I	3 500
	Kpn I	3 900
	Eco RI	4 100
5	Eco RI	5 300
	Sal I	5 500
	Kpn I	6 100
	Bgl II	6 500
	Bgl II	7 600
10	Hind III	7 850
	Bam HI	8 150
	Xho I	8 600
	Kpn I	8 700
	Bgl II	8 750
15	Bgl II	9 150
	Sac I	9 200
	Hind III	9 250

Another DNA variant according to this invention optionally contains an additional Hind III approximately
20 at the 5 550 coordinate.

Reference is further made to fig. 1 which shows a more detailed restriction map of said whole-DNA (λ J19).

An even more detailed nucleotide sequence of a preferred DNA according to the invention is shown in fig.
25 4-12 hereafter.

The invention further relates to other preferred DNA fragments which will be referred to hereafter.

Additional features of the invention will appear in the course of the non-limitative disclosure of additional features of preferred DNAs of the invention, as well
30 as of preferred polypeptides according to the invention. Reference will further be had to the drawings in which :
- fig. 1 is the restriction map of a complete LAV genome (clone λ J19) ;

35 - figs. 2 and 3 show diagrammatically parts of the three

possible reading phases of LAV genomic RNA, including the open reading frames (ORF) apparent in each of said reading phases :

- figs. 4-12 show the successive nucleotidic sequences of a complete LAV genome. The possible peptide sequences in relation to the three possible reading phases related to the nucleotidic sequences shown are also indicated ;
- figs. 13-18 reiterate the sequence of part of the LAV genome containing the genes coding for the envelope proteins, with particular boxed peptidic sequences which corresponds to groups which normally carry glycosyl groups.

The sequencing and determination of sites of particular interest was carried out on a phage recombinant corresponding to AJ19 disclosed in the abovesaid British Patent application Nr. 84 23659. A method for preparing it is disclosed in that application.

The whole recombinant phage DNA of clone AJ19 (disclosed in the earlier application) was sonicated according to the protocol of DEININGER (1983), Analytical Biochem. 129, 216. the DNA was repaired by a Klenow reaction for 12 hours at 16°C. The DNA was electrophoresed through 0.8 % agarose gel and DNA in the size range of 300-600 bp was cut out and electroeluted and precipitated. Resuspended DNA (in 10 mM Tris, pH 8 ; 0.1 mM EDTA) was ligated into M13mp8 RF DNA (cut by the restriction enzyme SmaI and subsequently alkaline phosphated), using T4 DNA- and RNA-ligases (Maniatis T et al (1982) - Molecular cloning - Cold Spring Harbor Laboratory). An *E. coli* strain designated as TGI was used for further study. This strain has the following genotype :

Alac pro, supE, thi.F'traD36, proA8, lacI^q, ZAM15,r⁻

This *E. coli* TGI strain has the peculiarity of enabling recombinants to be recognized easily. The blue colour of the cells transfected with plasmids which did

not recombine with a fragment of LAV DNA is not modified. To the contrary cells transfected by a recombinant plasmid containing a LAV DNA fragment yield white colonies. The technique which was used is disclosed in Gene (1983), 26.

5 101.

This strain was transformed with the ligation mix using the Hanahan method (Hanahan D (1983) J. Mol. Biol. 168, 557). Cells were plated out on tryptone-agarose plate with IPTG and X-gal in soft agarose. White plaques were
10 either picked and screened or screened directly in situ using nitrocellulose filters. Their DNAs were hybridized with nick-translated DNA inserts of pUC18 Hind III subclones of AJ19. this permitted the isolation of the plasmids or subclones of λ which are identified in the
15 table hereafter. In relation to this table it should also be noted that the designation of each plasmid is followed by the deposition number of a cell culture of E. coli TGI containing the corresponding plasmid at the "Collection Nationale des Cultures de Micro-organismes" (C.N.C.M.) of
20 the Pasteur Institute in Paris, France. A non-transformed TGI cell line was also deposited at the C.N.C.M. under Nr. I-364. All these deposits took place on November 15, 1984. The sizes of the corresponding inserts derived from the LAV genome have also been indicated.

25



TABLE

Essential features of the recombinant plasmids

5	- pJ19 - 1 plasmid	(I-365)	0.5 kb
	Hind III - Sac I - Hind III		
	- pJ19 - 17 plasmid	(I-367)	0.6 kb
10	Hind III - Pst I - Hind III		
	- pJ19 - 6 plasmid	(I-368)	1.5 kb
15	Hind III (5')		
	Sam HI		
	Xho I		
	Kpn I		
	Bgl II		
20	Sac I (3')		
	Hind III		
	- pJ19-13 plasmid	(I-368)	6.7 kb
25	Hind III (5')		
	Bgl II		
	Kpn I		
	Kpn I		
	Eco RI		
30	Eco RI		
	Sal I		
	Kpn I		
	Bgl II		
	Bgl II		
35	Hind III (3')		

Positively hybridizing M13 phage plates were grown up for 5 hours and the single-stranded DNAs were extracted.

M13mp8 subclones of λ J19 DNAs were sequenced according to the dideoxy method and technology devised by Sanger et al (Sanger et al (1977), Proc. Natl. Acad. Sci. USA. 74, 5483 and M13 cloning and sequencing handbook, AMERSHAM (1983). the 17-mer oligonucleotide primer α -³⁵SdATP (400Ci/mmol, AMERSHAM), and 0.5X-5X buffer gradient gels (Biggen M.D. et al (1983, Proc. Natl. Acad. Sci. USA. 50, 3963) were used. Gels were read and put into the computer under the programs of Staden (Staden R. (1982), Nucl. Acids Res. 10, 4731). All the appropriate references and methods can be found in the AMERSHAM M13 cloning and sequencing handbook.

The complete sequence of λ J19 was deduced from the experiments as further disclosed hereafter.

Figs. 4-12 provide the DNA nucleotide sequence of the complete genome of LAV. The numbering of the nucleotides starts from a left most Hind III restriction site (5'AAG...) of the restriction map. The numbering occurs in tens whereby the last zero number of each of the numbers occurring on the drawings is located just below the nucleotide corresponding to the nucleotides designated. I.e. the nucleotide at position 10 is T, the nucleotide at position 20 is C, etc..

Above each of the lines of the successive nucleotidic sequences there are provided three lines of single letters corresponding to the aminoacid sequence deduced from the DNA sequence (using the genetic code) for each at the three reading phases, whereby said single letters have the following meanings.

A : alanine
R : arginine
K : lysine
H : histidine
C : cysteine

M : méthionine
 W : tryptophan
 F : phenylalanine
 Y : tyrosine
 5 L : leucine
 V : valine
 I : isoleucine
 G : glycine
 T : thréonine
 10 S : serine
 E : glutamic acid
 D : Aspartic acid
 N : asparagine
 Q : glutamine
 15 P : proline.

The asterik signs "*" correspond to stop codons (i.e. TAA, TAG and TGA).

Starting above the first line of the DNA nucleotidic sequence of fig. 4 the three reading phases are respectively marked "1", "2", "3", on the left handside of the drawing. The same relative presentation of the three theoretical reading phases is then used all over the successive lines of the LAV nucleotidic sequence.

Figs. 2 and 3 provide a diagrammatized representation of the lengths of the successive open reading frames corresponding to the successive reading phases (also referred to by numbers "1", "2" and "3" appearing in the left handside part of fig. 2). The relative positions of these open reading frames (ORF) with respect to the nucleotidic structure of the LAV genome is referred to by the scale of numbers representative of the respective positions of the corresponding nucleotides in the DNA sequence. The vertical bars correspond to the positions of the corresponding stop codons.

35 1) The "gag gene" (or ORF-gag)

The "gag gene" codes for core proteins.

Particularly it appears that a genomic fragment (ORF-gag) thought to code for the core antigens including the p25, p18 and p13 proteins is located between nucleotidic position 236 (starting with 5' CTA GCG GAG 3') and
 5 nucleotidic position 1759 (ending by CTCG TCA CAA 3'). The structure of the peptides or proteins encoded by parts of said ORF is deemed to be that corresponding to phase 2.

The methionine aminoacid "M" coded by the ATG at position 260-262 is the probable initiation methionine of
 10 the gag protein precursor. The end of ORF-gag and accordingly of gag protein appears to be located at position 1759.

The beginning of p25 protein, thought to start by a P-I-V-Q-N-I-Q-G-Q-M-V-H aminoacid sequence is
 15 thought to be coded for by the nucleotidic sequence CCTATA..., starting at position 858.

Hydrophilic peptides in the gag open reading frame are identified hereafter. They are defined starting from aminoacid 1 = Met (M) coded by the ATG starting from 260-2
 20 in the LAV DNA sequence.

Those hydrophilic peptides are
 12-32 aminoacids inclusive

	37-46	-	-
	49-79	-	-
25	88-153	-	-
	158-185	-	-
	178-188	-	-
	200-220	-	-
	226-234	-	-
30	239-264	-	-
	288-331	-	-
	352-381	-	-
	377-390	-	-
	399-432	-	-
35	437-484	-	-
	492-498	-	-

The invention also relates to any combination of these peptides.

2) The "pol gene" (or ORF-pol)

Figs. 4-12 also show that the DNA fragments
 5 extending from nucleotidic position 1555 (starting with
 5' TTT TTT 3' to nucleotidic position 5086 is thought
 to correspond to the pol gene. The polypeptidic structure
 of the corresponding polypeptides is deemed to be that
 corresponding to phase 1. It stops at position 4583 (and
 10 by 5' G GAT GAG GAT 3').

These genes are thought to code for the virus
 polymerase or reverse transcriptase.

3) The envelope gene (or ORF-env)

The DNA sequence thought to code for envelope
 15 proteins is thought to extend from nucleotidic position
 5670 (starting with 5' AAA GAG GAG A.... 3') up to nucleo-
 tidic position 8132 (ending by A ACT AAA GAA 3').
 Polypeptidic structures of sequences of the envelope
 protein correspond to those read according to the "phase
 20 3" reading phase.

The start of env transcription is thought to be at
 the level of the ATG codon at positions 5691-5693.

Additional feature of the envelope protein coded
 by the env genes appear on figs. 13-18. These are to be
 25 considered as paired figs. 13 and 14 ; 15 and 16 ; 17 and
 18 respectively.

It is to be mentioned that because of format
 difficulties.

Fig. 14 overlaps to some extent with fig. 13.

30 Fig. 16 overlaps to some extent with fig. 15.

Fig. 18 overlaps to some extent with fig. 17.

Thus for instance figs. 13 and 14 must be con-
 sidered together. Particularly the sequence shown on the
 first line on the top of fig. 13 overlaps with the
 35 sequence shown on the first line on the top of fig. 14. In
 other words the starting of the reading of the successive

sequences of the env gene as represented in figs. 13-18 involves first reading the first line at the top of fig. 13 then proceeding further with the first line of fig. 14. One then returns to the beginning of the second line of fig. 13, then again further proceed with the reading of the second line of page 14, etc... The same observations then apply to the reading of the paired figs. 15 and 16, and paired figs. 17 and 18, respectively.

The locations of neutralizing epitopes are further apparent in figs. 13-18. reference is more particularly made to the boxed groups of three letters included in the aminoacid sequences of the envelope proteins (reading phase 3) which can be designated generally by the formula N-X-S or N-X-T, wherein X is any other possible aminoacid. Thus the initial protein product of the env gene in a glycoprotein of molecular weight in excess of 91,000. These groups are deemed to generally carry glycosylated groups. These N-X-S and N-X-T groups with attached glycosylated groups form together hydrophylic regions of the protein and are deemed to be located at the periphery of and to be exposed outwardly with respect to the normal conformation of the proteins. Consequently they are considered as being epitopes which can efficiently be brought into play in vaccine compositions.

The invention thus concerns with more particularity peptide sequences included in the env-proteins and excizable therefrom (or having the same aminoacid structure), having sizes not exceeding 200 aminoacids.

Preferred peptides of this invention (referred to hereafter as a, b, c, d, e, f) are deemed to correspond to those encoded by the nucleotide sequences which extend respectively between the following positions :

- a) from about 8095 to about 8200 ✓
- b) " " 8260 " " 8310 ✓
- 35 c) " " 8390 " " 8440 ✓
- d) " " 8485 " " 8620 ✓

e) - - 6860 - - 5930 ✓
 f) - - 7535 - - 7630 ✓

Other hydrophilic peptides in the env open reading frame are identified hereafter. they are defined starting from

aminoacid 1 = lysine (K) coded by the AAA at position 5670-2 in the LAV DNA sequence.

These hydrophilic peptides are

8-23 aminoacids inclusive

18	63-78	-	-
	82-90	-	-
	97-123	-	-
	127-183	-	-
	197-201	-	-
15	239-294	-	-
	300-327	-	-
	334-381	-	-
	397-424	-	-
	488-500	-	-
20	510-523	-	-
	551-577	-	-
	594-603	-	-
	621-630	-	-
	657-679	-	-
25	719-758	-	-
	780-803	-	-

The invention also relates to any combination of these peptides.

4) The other ORF

The invention further concerns DNA sequences which provide open reading frames defined as ORF-Q, ORF-R and as "1", "2", "3", "4", "5", the relative position of which appears more particularly in figs. 2 and 3.

These ORFs have the following locations :

35	ORF-Q	phase 1	start 4478	stop 5086
	ORF-R	" 2	" 8249	" 8896

ORF-1	-	1	-	5029	-	5316
ORF-2	-	2	-	5273	-	5515
ORF-3	-	1	-	5383	-	5616
ORF-4	-	2	-	5519	-	5773
5 ORF-5	-	1	-	7966	-	8279

The LTR (long terminal repeats) can be defined as lying between position 8560 and position 160 (end extending over position 9097/1). As a matter of fact the end of the genome is at 9097 and, because of the LTR structure of the retrovirus, links up with the beginning of the sequence :

Hind III
CTCAATAAAGCTTGCCTTG

15

9097 1

The invention concerns more particularly all the DNA fragments which have been more specifically referred to hereabove and which correspond to open reading frames. It will be understood that the man skilled in the art will be able to obtain them all, for instance by cleaving an entire DNA corresponding to the complete genome of a LAV species, such as by cleavage by a partial or complete digestion thereof with a suitable restriction enzyme and by the subsequent recovery of the relevant fragments. The different DNAs disclosed in the earlier mentioned British Application can be resorted to also as a source of suitable fragments. The techniques disclosed hereabove for the isolation of the fragments which were then included in the plasmids referred to hereabove and which were then used for the DNA sequencing can be used.

Of course other methods can be used. Some of them have been exemplified in the earlier British Application. reference is for instance made to the following methods.

a) DNA can be transfected into mammalian cells with appropriate selection markers by a variety of techniques, calcium phosphate precipitation, polyethylene

glycol, protoplast-fusion, etc..

b) DNA fragments corresponding to genes can be cloned into expression vectors for E. coli, yeast- or mammalian cells and the resultant proteins purified.

5 c) The proviral DNA can be "shot-gunned" (fragmented) into procaryotic expression vectors to generate fusion polypeptides. Recombinant producing antigenically competent fusion proteins can be identified by simply screening the recombinants with antibodies against LAV
10 antigens .

The invention also relates more specifically to cloned probes which can be made starting from any DNA fragment according to this invention, thus to recombinant DNAs containing such fragments, particularly any plasmids
15 amplifiable in procaryotic or eucaryotic cells and carrying said fragments.

Using the cloned DNA fragments as a molecular hybridization probe - either by marking with radionucleotides or with fluorescent reagents - LAV virion RNA may be
20 detected directly in the blood, body fluids and blood products (e.g. of the antihemophylic factors such as Factor VIII concentrates) and vaccines, i.e. hepatitis B vaccine. It has already been shown that whole virus can be detected in culture supernatants of LAV producing cells. A
25 suitable method for achieving that detection comprises immobilizing virus onto said a support e.g. nitrocellulose filters, etc., disrupting the virion and hybridizing with labelled (radiolabelled or "cold" fluorescent- or enzyme-labelled) probes. Such an approach has already been
30 developed for Hepatitis B virus in peripheral blood (according to SCOTTO J. et al. Hepatology (1983), 3, 379-384).

Probes according to the invention can also be used for rapid screening of genomic DNA derived from the tissue
35 of patients with LAV related symptoms, to see if the proviral DNA or RNA is present in host tissue and other

tissues.

A method which can be used for such screening comprise the following steps : extraction of DNA from tissue, restriction enzyme cleavage of said DNA, electrophoresis of the fragments and Southern blotting of genomic DNA from tissues, subsequent hybridization with labelled cloned LAV proviral DNA. Hybridization in situ can also be used.

Lymphatic fluids and tissues and other non-lymphatic tissues of humans, primates and other mammalian species can also be screened to see if other evolutionary related retrovirus exist. The methods referred to hereabove can be used, although hybridization and washings would be done under non stringent conditions.

The DNA according to the invention can be used also for achieving the expression of LAV viral antigens for diagnostic purposes.

The invention also relates to the polypeptides themselves which can be expressed by the different DNAs of the inventions, particularly by the ORFs or fragments thereof, in appropriate hosts, particularly procaryotic or eucaryotic hosts, after transformation thereof with a suitable vector previously modified by the corresponding DNAs.

These polypeptides can be used as diagnostic tools, particularly for the detection of antibodies in biological media, particularly in sera or tissues of persons afflicted with pre-AIDS or AIDS, or simply carrying antibodies in the absence of any apparent disorders. Conversely the different peptides according to this invention can be used themselves for the production of antibodies, preferably monoclonal antibodies specific of the different peptides respectively. For the production of hybridomas secreting said monoclonal antibodies conventional production and screening methods are used. These monoclonal antibodies, which themselves are part of

the invention then provide very useful tools for the identification and even determination of relative proportions of the different polypeptides or proteins in biological samples, particularly human samples containing
5 LAV or related viruses.

Thus all of the above peptides can be used in diagnostics as sources of immunogens or antigens free of viral particles, produced using non-permissive systems, and thus of little or no biohazard risk.

10 The invention further relates to the hosts (proca-ryotic or eucaryotic cells) which are transformed by the above mentioned recombinants and which are capable of expressing said DNA fragments.

Finally it also relates to vaccine compositions
15 whose active principle is to be constituted by any of the expressed antigens, i.e. whole antigens, fusion polypep-tides or oligopeptides in association with a suitable pharmaceutical or physiologically acceptable carrier.

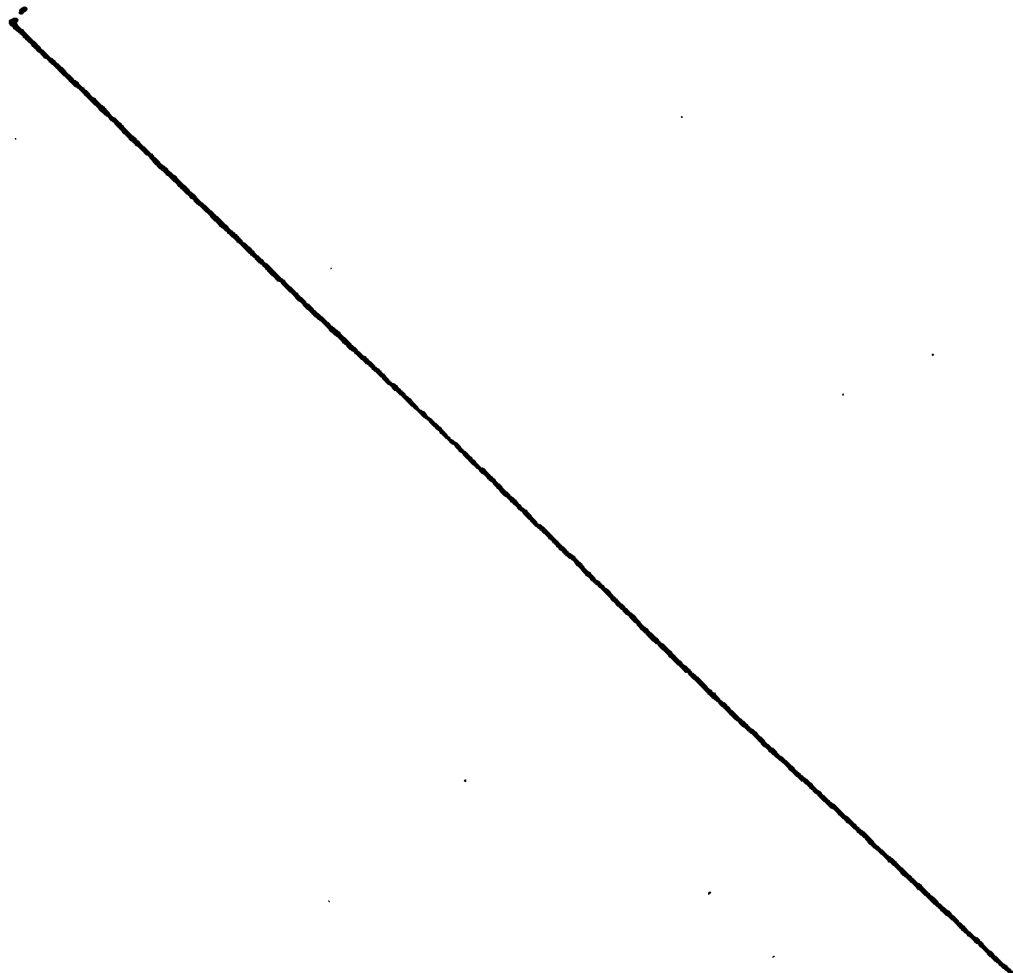
Preferably the active principles to be considered
20 in that field consist of the peptides containing less than 250 aminoacid units, preferably less than 150 as deducible for the complete genomes of LAV, and even more preferably those peptides which contain one or more groups selected from N-X-S and N-X-T as defined above. Preferred peptides
25 for use in the production of vaccinating principles are peptides (a) to (f) as defined above. By way of example having no limitative character, there may be mentioned that suitable dosages of the vaccine compositions are those which enable administration to the host,
30 particularly human host ranging from 10 to 500 micrograms per kg, for instance 50 to 100 micrograms per kg.

For the purpose of clarity figs. 19 to 26 are added. reference may be made thereto in case of difficul-ties of reading blurred parts of figs. 4 to 12.

Needless to say that figs. 19-25 are merely a reiteration of the whole DNA sequence of the LAV genome.

Finally the invention also concerns vectors for the transformation of eucaryotic cells of human origin, particularly lymphocytes, the polymerases of which are capable of recognizing the LTRs of LAV. Particularly said vectors are characterized by the presence of a LAV LTR therein, said LTR being then active as a promoter enabling the efficient transcription and translation in a suitable host of the above defined, of a DNA insert coding for a determined protein placed under its controls.

Needless to say that the invention extends to all variants of genomes and corresponding DNA fragments (ORFs) having substantially equivalent properties, all of said genomes belonging to retroviruses which can be considered as equivalents of LAV.



CLAIMS

1. A DNA fragment of LAV extending from nucleotide position 236 to nucleotide position 1759.

2. A DNA fragment of LAV extending from nucleotide position 1535 to nucleotide position 5086.

3. A DNA fragment of LAV extending from nucleotide position 5670 to nucleotide position 8132.

4. A vector containing a DNA fragment according to any of claims 1 to 3.

5. Peptide corresponding to any of those encoded by the nucleotide sequences which extend respectively between the following positions :

a) from about 6095 to about 6200.

b) " " 6260 " " 6310

c) " " 6390 " " 6440

d) " " 6485 " " 6520

e) " " 6860 " " 6930

f) " " 7535 " " 7630

6. Peptide characterized by a sequence of amino-acids deducible from LAV DNA the terminal aminoacids of which extend between the following positions with respect to the lysine (position 1) coded by the AAA at position 5670-5672 in the LAV DNA.

8-23 aminoacids inclusive

63-78 " "

82-90 " "

97-123 " "

127-183 " "

197-201 " "

239-294 " "

300-327 " "

334-381 " "

397-424 " "

466-500 " "

510-523 " "

551-577 " "

594-803 - -
 821-830 - -
 857-878 - -
 719-758 - -
 5 780-803 - -

or any combination of these peptides.

7. Peptide corresponding to the aminoacid sequences deducible from LAV DNA and the terminal aminoacids of which are positionned at the positions hereafter counted from the Met at position 1 coded by the ATG sequence at nucleotide positions 280-2 :

12-32 aminoacids inclusive

	37-48	-	-
	49-79	-	-
15	88-153	-	-
	158-165	-	-
	178-188	-	-
	200-220	-	-
	228-234	-	-
20	239-264	-	-
	288-331	-	-
	352-381	-	-
	377-390	-	-
	399-432	-	-
25	437-484	-	-
	482-498	-	-

and combination of said peptides.

8. Diagnostic means containing any of the DNA fragments of any of claims 1 to 3.

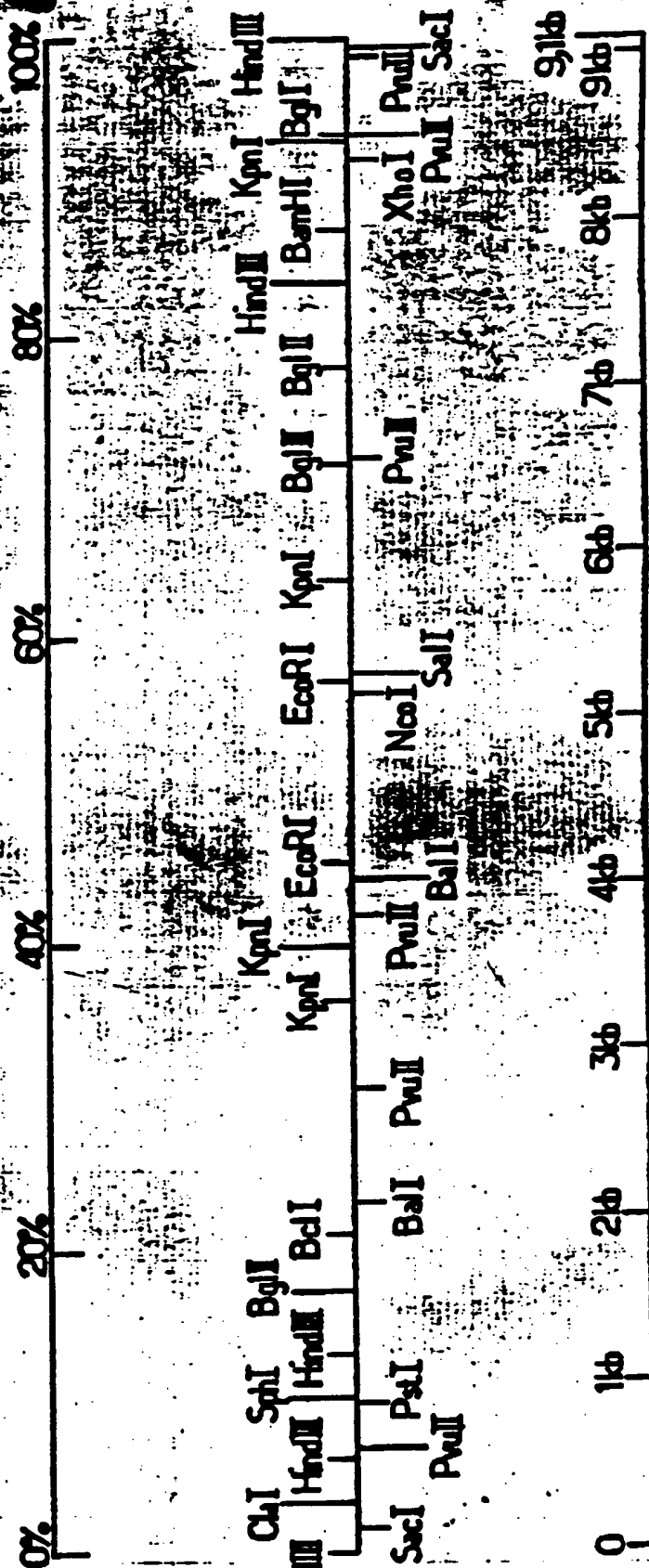
30 9. Diagnostic means containing any of the peptides of any of claims 4 to 8.

10. Vaccine compositions containing any of the peptides according to any of claims 4 to 8 in association with a pharmaceutical vehicle.

35

End of transmission

FIG.1.



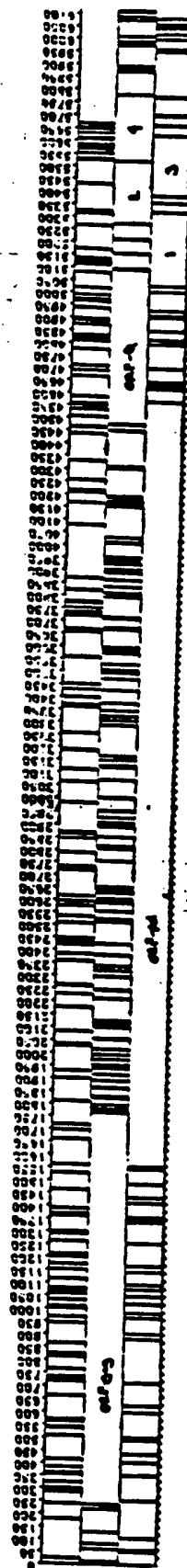


Fig. 2

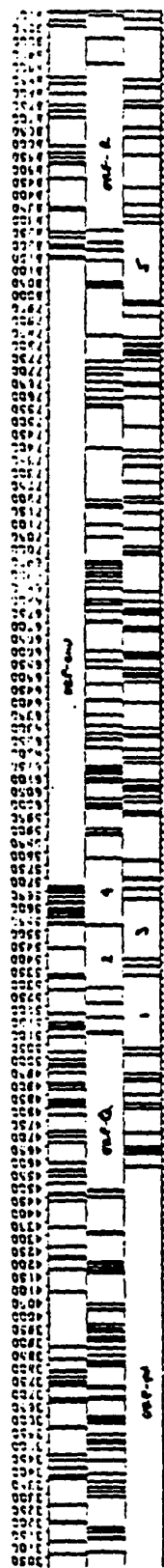


Fig. 3

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

10-1 C

[illegible][illegible][illegible][illegible][illegible][illegible]

Fig 4

K 01 V O P V J M S G H M I P I M H I L O D L C O P V L O H S M A S F I C C K
 a w Y S P I S I L U I O J C P K F P F O J V V E K F Y V I L Q A F O A S O I V N
 I C I A L P A F W T O M D U M A P L F I M O I G S I A L O F P S K L M B O O Y
 A G A A T A G C C C T A G C A T C T G C A L A T A A G C A C A C C C T T A G A C T A T G A G A C C G T C T C T A T A A A C T C T A G A G C C C A C C A C C T C A C A C G C A T A A
 1010 1170 1170 1120 1120 1130 1150 1150 1160 1170 1190 1190 1200

M L D D A V L V G P M C E P A L O D V F U S I G T S S V T R M D D S Y S G S C
 M A T E T L L V O M A H P U R K T I L K A L G P A A T L E E P M T A C O C V C
 I G O J M P C A S M N N T Q I V R L F O K M W D G C L M O K O O H V R E W E
 A A T T G A C A C A A A C T T T G C C A A A T G C C A A T T T T A A G A C T T G G A C C A C C A C T A C T A C A G A A T A T C A C A G A C T A C A C C G C A C T G G
 1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320

[illegible]

L M U A R A M S G K L U G P O E K G L L F A A K G A T P V F P L V O F T G O F F F
 C A K E G H I A A U C A P L G M K G C M C G M O M K I T E R O A M F L L
 T T G C G A A C A A G C C A T A G C C A G A A A T T C A G C C A A T T C A G C A C A C C G A A T T T T T
 1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560 1570 1580 1590

4 E D L A F L D G A A Z F S S E U T R A Y S P T A Q E L O V A G A D N M Y L S
C A I A P S Y V M C N P C N I L S V P E S T A P P E S F R S C V E T T I P S O
G A G G L P T O E C C C I F A C A O U S U O P C C K R A S C L C O R O O L P L N
A G G A G A T C G C C T T C T A A G C A G C C C A T T T T C A G C A C C A G C C A C C C C C C A G C T T C A G C T C T C G G T A G A C A G A C T C T C T C A
1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680

[illegible]

6 A 2 0 1 1 V L F E 7 3 L P C M 4 P 4 4 1 2 6 1 6 6 F 1 M V 1 0 V 0 1 L 1 E 1
F C 4 1 0 V 0 4 M 0 J C J F 10 F 4 0 4 0 C F L E V L S M 0 0 S V 1 0 V S 0 4 S
S Q 4 5 1 4 4 V F 1 A 7 P E 1 4 4 0 0 C M W F V Q S 4 7 V 0 1 0 F 1 H 4 M L
S C 6 C A G C T A C G C T A T A C A G C A A A C G C A A A A A T C A C C C C C A A A T C A G G C T T T A T C A A G C A A G A C A T A T A T A C C A C A C A C A C A A G A A T C
1470 1480 1490 1500 1510 1520 1530 1540 1550 1560 1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680 1690 1700 1710 1720

[illegible]

Fig. 5

[illegible]

Nov 27 11 44 AM '67

[illegible][illegible][illegible][illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible][illegible][illegible]

[illegible]

Q. 12

[illegible]

2210
 2220
 2230
 2240
 2250
 2260
 2270
 2280
 2290
 2300
 2310
 2320
 2330
 2340
 2350
 2360
 2370
 2380
 2390
 2400
 2410
 2420
 2430
 2440
 2450
 2460
 2470
 2480
 2490
 2500
 2510
 2520
 2530
 2540
 2550
 2560
 2570
 2580
 2590
 2600
 2610
 2620
 2630
 2640
 2650
 2660
 2670
 2680
 2690
 2700
 2710
 2720
 2730
 2740
 2750
 2760
 2770
 2780
 2790
 2800
 2810
 2820
 2830
 2840
 2850
 2860
 2870
 2880
 2890
 2900
 2910
 2920
 2930
 2940
 2950
 2960
 2970
 2980
 2990
 3000
 3010
 3020
 3030
 3040
 3050
 3060
 3070
 3080
 3090
 3100
 3110
 3120
 3130
 3140
 3150
 3160
 3170
 3180
 3190
 3200
 3210
 3220
 3230
 3240
 3250
 3260
 3270
 3280
 3290
 3300
 3310
 3320
 3330
 3340
 3350
 3360
 3370
 3380
 3390
 3400
 3410
 3420
 3430
 3440
 3450
 3460
 3470
 3480
 3490
 3500
 3510
 3520
 3530
 3540
 3550
 3560
 3570
 3580
 3590
 3600
 3610
 3620
 3630
 3640
 3650
 3660
 3670
 3680
 3690
 3700
 3710
 3720
 3730
 3740
 3750
 3760
 3770
 3780
 3790
 3800
 3810
 3820
 3830
 3840
 3850
 3860
 3870
 3880
 3890
 3900
 3910
 3920
 3930
 3940
 3950
 3960
 3970
 3980
 3990
 4000
 4010
 4020
 4030
 4040
 4050
 4060
 4070
 4080
 4090
 4100
 4110
 4120
 4130
 4140
 4150
 4160
 4170
 4180
 4190
 4200
 4210
 4220
 4230
 4240
 4250
 4260
 4270
 4280
 4290
 4300
 4310
 4320
 4330
 4340
 4350
 4360
 4370
 4380
 4390
 4400
 4410
 4420
 4430
 4440
 4450
 4460
 4470
 4480
 4490
 4500
 4510
 4520
 4530
 4540
 4550
 4560
 4570
 4580
 4590
 4600
 4610
 4620
 4630
 4640
 4650
 4660
 4670
 4680
 4690
 4700
 4710
 4720
 4730
 4740
 4750
 4760
 4770
 4780
 4790
 4800
 4810
 4820
 4830
 4840
 4850
 4860
 4870
 4880
 4890
 4900
 4910
 4920
 4930
 4940
 4950
 4960
 4970
 4980
 4990
 5000
 5010
 5020
 5030
 5040
 5050
 5060
 5070
 5080
 5090
 5100
 5110
 5120
 5130
 5140
 5150
 5160
 5170
 5180
 5190
 5200
 5210
 5220
 5230
 5240
 5250
 5260
 5270
 5280
 5290
 5300
 5310
 5320
 5330
 5340
 5350
 5360
 5370
 5380
 5390
 5400
 5410
 5420
 5430
 5440
 5450
 5460
 5470
 5480
 5490
 5500
 5510
 5520
 5530
 5540
 5550
 5560
 5570
 5580
 5590
 5600
 5610
 5620
 5630
 5640
 5650
 5660
 5670
 5680
 5690
 5700
 5710
 5720
 5730
 5740
 5750
 5760
 5770
 5780
 5790
 5800
 5810
 5820
 5830
 5840
 5850
 5860
 5870
 5880
 5890
 5900
 5910
 5920
 5930
 5940
 5950
 5960
 5970
 5980
 5990
 6000
 6010
 6020
 6030
 6040
 6050
 6060
 6070
 6080
 6090
 6100
 6110
 6120
 6130
 6140
 6150
 6160
 6170
 6180
 6190
 6200
 6210
 6220
 6230
 6240
 6250
 6260
 6270
 6280
 6290
 6300
 6310
 6320
 6330
 6340
 6350
 6360
 6370
 6380
 6390
 6400
 6410
 6420
 6430
 6440
 6450
 6460
 6470
 6480
 6490
 6500
 6510
 6520
 6530
 6540
 6550
 6560
 6570
 6580
 6590
 6600
 6610
 6620
 6630
 6640
 6650
 6660
 6670
 6680
 6690
 6700
 6710
 6720
 6730
 6740
 6750
 6760
 6770
 6780
 6790
 6800
 6810
 6820
 6830
 6840
 6850
 6860
 6870
 6880
 6890
 6900
 6910
 6920
 6930
 6940
 6950
 6960
 6970
 6980
 6990
 7000
 7010
 7020
 7030
 7040
 7050
 7060
 7070
 7080
 7090
 7100
 7110
 7120
 7130
 7140
 7150
 7160
 7170
 7180
 7190
 7200
 7210
 7220
 7230
 7240
 7250
 7260
 7270
 7280
 7290
 7300
 7310
 7320
 7330
 7340
 7350
 7360
 7370
 7380
 7390
 7400
 7410
 7420
 7430
 7440
 7450
 7460
 7470
 7480
 7490
 7500
 7510
 7520
 7530
 7540
 7550
 7560
 7570
 7580
 7590
 7600
 7610
 7620
 7630
 7640
 7650
 7660
 7670
 7680
 7690
 7700
 7710
 7720
 7730
 7740
 7750
 7760
 7770
 7780
 7790
 7800
 7810
 7820
 7830
 7840
 7850
 7860
 7870
 7880
 7890
 7900
 7910
 7920
 7930
 7940
 7950
 7960
 7970
 7980
 7990
 8000
 8010
 8020
 8030
 8040
 8050
 8060
 8070
 8080
 8090
 8100
 8110
 8120
 8130
 8140
 8150
 8160
 8170
 8180
 8190
 8200
 8210
 8220
 8230
 8240
 8250
 8260
 8270
 8280
 8290
 8300
 8310
 8320
 8330
 8340
 8350
 8360
 8370
 8380
 8390
 8400
 8410
 8420
 8430
 8440
 8450
 8460
 8470
 8480
 8490
 8500
 8510
 8520
 8530
 8540
 8550
 8560
 8570
 8580
 8590
 8600
 8610
 8620
 8630
 8640
 8650
 8660
 8670
 8680
 8690
 8700
 8710
 8720
 8730
 8740
 8750
 8760
 8770
 8780
 8790
 8800
 8810
 8820
 8830
 8840
 8850
 8860
 8870
 8880
 8890
 8900
 8910
 8920
 8930
 8940
 8950
 8960
 8970
 8980
 8990
 9000
 9010
 9020
 9030
 9040
 9050
 9060
 9070
 9080
 9090
 9100
 9110
 9120
 9130
 9140
 9150
 9160
 9170
 9180
 9190
 9200
 9210
 9220
 9230
 9240
 9250
 9260
 9270
 9280
 9290
 9300
 9310
 9320
 9330
 9340
 9350
 9360
 9370
 9380
 9390
 9400
 9410
 9420
 9430
 9440
 9450
 9460
 9470
 9480
 9490
 9500
 9510
 9520
 9530
 9540
 9550
 9560
 9570
 9580
 9590
 9600
 9610
 9620
 9630
 9640
 9650
 9660
 9670
 9680
 9690
 9700
 9710
 9720
 9730
 9740
 9750
 9760
 9770
 9780
 9790
 9800
 9810
 9820
 9830
 9840
 9850
 9860
 9870
 9880
 9890
 9900
 9910
 9920
 9930
 9940
 9950
 9960
 9970
 9980
 9990
 10000

Fig 11

Start ORF - R

000-
C C A W V K N O C G H A Y E K G M Y F I S O A S B A C C S I S R P C K T
V A S C O M V D L O C P L O G M E R R A F P A A O C V C A S S A S O L F M H C
GGTGGCAATGTCACAAAGCTATGTCTTGCGATCCCTACTGAAGGGAAGAATGAGAACAGTCCACGCACATCGGCGGAGCACATCTTCAGACGCCAAAATCGG
8290 4300 4310 4320 4330 4340 4350 4360 4370 4380 4390 4400

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

K A A V U L S H F L M E K C C L E C L I H S O P E M I L D L M I V Y I O C V F
 R Q L O I L A T F O K M R C D M K C O F P M E D M I S L I C G S Y T M A T S
 CAAGCAGCTGTAGTCTAGCCATTTTAAAGAAAGCCGCGATCGACGGCTATTCTATCTCCACGACGACATCTCTGATCTACCCACACAGCCTATT
 8530 8540 8550 8560 8570 8580 8590 8600 8610 8620 8630 8640

P O L A E L M I R A A G O I S T D L M N V L O A S T S O A R O C R R C O O R P E M
 P D M Q N V T P C P C V A P L T F C M C V M L V P V E P D M V E A M N C E M
 L I C M N I M J C U C S D I H O P L D C A I S V O Q L S O I R O A R P I N E R
 C C T G A T T G C C A A C A C C C G G C T C A G A T A C C A C C T T T G A T G C T T A A G C T A T A C C A G T G A C C C A T A A G C G C C A T A A G C G A C A
 8630 8660 8670 8680 8690 8700 8710 8720 8730 8740 8750 8760

[illegible][illegible][illegible]

12

9097

fig 12